Fermi National Accelerator Laboratory

The Stromlo- APM Redshift Survey IV.
The Redshift Catalogue

J. Loveday et al.

Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510

March 1996

Disclaimer

The Stromlo-APM Redshift Survey IV.
The Redshift Catalogue

J. Loveday

Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA; loveday@fnal.gov

B.A. Peterson

Mount Stromlo and Siding Spring Observatories, Weston Creek PO, ACT 2611, Australia;
peterson@mso.anu.edu.au

S.J. Maddox

Royal Greenwich Observatory, Madingley Road, Cambridge, CB3 0EZ, England;
sjm@mail.ast.cam.ac.uk

and

G. Efstathiou

Department of Physics, Keble Road, Oxford, OX1 3RH, England;
g.efstathiou@physics.oxford.ac.uk

ABSTRACT

The Stromlo-APM Redshift Survey consists of 1797 galaxies with $b_J \leq 17.15$ selected randomly at a rate of 1 in 20 from Automated Plate Measurement (APM) scans. The survey covers a solid angle of 1.3 sr (4300 square degrees) in the south galactic cap. Redshifts have been measured for 1790 (99.6%) of the galaxies. The median galaxy recession velocity is 15,300 km/s and so the volume probed is $V \approx 1.38 \times 10^6 \ h^{-3}\mathrm{Mpc}^3$. In this paper, we describe the construction of the redshift catalogue and present the survey data.

Subject headings: galaxies: distances and redshifts — galaxies: fundamental parameters — large-scale structure of universe — methods: observational — techniques: radial velocities — surveys

1. Introduction

Galaxy surveys are of fundamental importance for testing the predictions of competing theories for the formation of galaxies and large scale structure. Measurement of the galaxy

luminosity function over a wide range of magnitudes and environments provides important constraints on models of galaxy formation. Comparison of observed galaxy clustering with large scale structure models is most robust on large scales, where linear perturbation theory can be used. It is thus very desirable to accurately estimate the galaxy correlation function $\xi(r)$ on scales where $\xi(r)$ is much less than unity, *i.e.* $r \gtrsim 20h^{-1}\mathrm{Mpc}$. There is then very strong motivation to survey as large a volume of the nearby Universe as possible.

When the Stromlo-APM survey was started in 1987, the most extensive redshift sample available was the Center for Astrophysics (CfA) redshift survey of Huchra *et al.* (1983) to $m_B = 14.5$, along with the first CfA 'slice' (de Lapparent, Geller and Huchra 1986) to $m_B = 15.5$. The CfA slice is characterized by large voids of diameter up to $\sim 50h^{-1}\mathrm{Mpc}$ surrounded by galaxies apparently distributed in thin sheets. By considering the distribution of voids in the slice, de Lapparent, Geller and Huchra (1988) estimated that the mean density of galaxies in the slice was uncertain by 25% and that for the $m_B = 14.5$ survey it was uncertain by $\sim 100\%$. Even now that the CfA survey has been extended to include 12 $6°$ slices containing over 9000 galaxies, the mean galaxy density is still uncertain by $\sim 25\%$ (Marzke *et al.* 1994).

Deep photometric surveys, such as the APM galaxy survey (Maddox *et al.* 1990a,b) sample a volume orders of magnitude larger than any existing redshift survey. Analysis of the APM survey by Maddox *et al.* (1990c) showed that the amplitude of the angular correlation function $w(\theta)$ on large scales was higher than that measured by Groth and Peebles (1977) for the Lick survey (Shane and Wirtanen 1967, Seldner *et al.* 1977) and that predicted by the 'standard' cold dark matter model (*e.g.* Davis *et al.* 1985, White *et al.* 1987). It was important to check the reliability of this result by directly calculating the spatial correlation function $\xi(r)$, as the angular correlation function $w(\theta)$ is extremely sensitive to small systematic effects such as plate matching errors and variable obscuration.

There was clearly a need then for a new redshift survey sampling a much larger volume than existing redshift surveys, yet a complete survey would have required prohibitively large amounts of telescope time. We therefore decided to undertake a randomly sampled *sparse* redshift survey, as advocated by Kaiser (1986).

In previous papers in this series, we have measured the large scale clustering of galaxies in redshift space (Loveday *et al.* 1992a), the galaxy luminosity function (Loveday *et al.* 1992b, hereafter Paper 1), the variation of galaxy clustering with luminosity and morphology (Loveday *et al.* 1995, Paper 2) and most recently we have estimated the combined cosmological density parameter and galaxy bias factor $\beta = \Omega_0^{0.6}/b$ from redshift-space distortions (Loveday *et al.* 1996, Paper 3). In this paper, we describe the construction of the redshift survey and present the catalogue data. The paper is organized as follows. We describe our sparse-sampling strategy in §2, and the construction of the source catalogue in §3. The observations are described in §4 and the reduction process is detailed in §5. Measurement of emission-line and cross-correlation velocities is described in §6. In §7 we present statistics on the measured redshifts and assess the accuracy

and reliability of the catalogue. The survey is summarized in §8.

2. Sparse Sampling Strategy

The main aim of the sparse sampling strategy for the Stromlo-APM survey was to minimize errors in the estimation of the spatial two-point correlation function $\xi(r)$ in the linear regime.

In order to ascertain the significance of the measured correlation function on large scales where the amplitude is low, it is important to have estimates of the statistical uncertainty. Since we know that galaxies are strongly clustered on small scales, Poissonian error estimates will underestimate the true errors. When $\xi(r')$ is small compared to unity, and when it is measured in a broad bin $r' \leq r \leq r' + \Delta r$, $\Delta r \approx r'$, then a realistic estimate of the uncertainty ε_ξ is given by

$$\varepsilon_\xi \approx \frac{1 + 4\pi n J_3}{\sqrt{N_p}}, \tag{1}$$

(Peebles 1973, Kaiser 1986), where

$$J_3(r') = \int_0^{r'} r^2 \xi(r) dr, \tag{2}$$

n is the galaxy number density and N_p the number of galaxy pairs for this distance bin. This result seems reasonable if one considers the 'cluster model' (Peebles 1980) which describes the distribution of galaxies as a randomly distributed set of clusters, each with $N_c = 1 + 4\pi n J_3$ members. The number of *independent* objects in this model is thus reduced by the factor $1/N_c$ and the statistical uncertainty in ξ increases accordingly.

If we sample a fraction f of galaxies at random, then the number density decreases to $f \times n$ and the number of pairs $N_p \propto f^2$. Thus for large N_c ($4\pi n J_3 \gg 1$), ε_ξ is almost independent of f. Kaiser (1986) has shown that the optimal sampling rate f (*i.e.* that which minimizes ε_ξ on large scales) is given roughly by $f \approx 7/[3(N_c - 1)]$. For the CfA redshift survey (Davis and Huchra 1982, Davis and Peebles 1983) we have $N_c \approx 50$ and so a sampling rate $f = 1/20$ was chosen for the Stromlo-APM survey. As well as being optimal for measurement of large scale structure, such a strategy is also extremely beneficial in estimating the luminosity function and mean density of galaxies (Paper 1).

3. The Source Catalogue

Our redshift survey is primarily based on the APM Bright Galaxy Catalogue (Loveday 1996, APM-BGC) which was compiled from 180 UK SERC J-survey Schmidt plates. The plates were scanned with the Automated Photographic Measurement (APM) system at Cambridge and cover 4180 square degrees of the southern sky approximately defined in equatorial coordinates

by $21^h \leqslant \alpha \leqslant 5^h$, $-72.5° \leqslant \delta \leqslant -17.5°$. To construct the APM Bright Galaxy Catalogue, every extended image with $b_J \leqslant 16.5$ was inspected by JL on film copies of the Schmidt plates and each galaxy assigned a morphological classification: elliptical, lenticular, spiral, irregular, uncertain or merged. The criterion for selecting extended images (as opposed to stellar images) was chosen conservatively to reject $\leqslant 5\%$ of compact, high surface brightness galaxies. Survey completeness at the bright end has been checked by comparison with the ESO catalogue (Lauberts 1982).

The visual inspection enables all non-galaxy images (about 70% of the candidates) to be rejected, thus ensuring that contamination by stars is negligible. This procedure is necessary for images brighter than $b_J \approx 16.5$ since halos around bright stellar images on Schmidt plates cause difficulties with star-galaxy classification using the measuring machine parameters alone. Photographic magnitudes have been converted to the total b_J system (Couch and Newell 1980) by comparison with CCD photometry obtained with the Mt. Stromlo and Siding Spring Observatories (MSSSO) $40''$ telescope between 1988 and 1990. Full details of star-galaxy separation, photometric matching and calibration for the APM Bright Galaxy Catalogue are described by Loveday (1996).

To form the redshift survey sample, galaxies were selected at random from the APM Bright Galaxy Catalogue at a rate of 1 in 20. To extend the redshift survey to fainter magnitudes, we selected objects at the same sampling rate from the deeper APM Galaxy Survey (Maddox *et al.* 1990a) that were brighter than $b_J = 17.15$ but not included in the Bright Galaxy Catalogue. We also included five additional survey fields: F362, F422, F423, F486 and F552. After visual inspection of these objects (by SJM) we ended up with a homogeneous random sample of 2011 galaxies.

Of these galaxy images, 214 were observed to be merged with other objects on the Schmidt plates, 70.5% of which are stars. Because it is difficult to measure magnitudes and redshifts for overlapping images, we have omitted these objects from our catalogue, leaving a sample of 1797 unmerged galaxies. Eliminating star-galaxy mergers may introduce a small gradient in the projected surface density of galaxies towards the galactic plane. In our clustering analyses (Papers 2 and 3), we minimized the effects of any such gradient by using Hamilton's (1993) estimator for $\xi(r)$. Eliminating galaxy-galaxy mergers can bias clustering measures even on large scales if high density regions are under-sampled, but since only 3% of galaxy images are classified as mergers, any such bias will be negligible.

Although our source catalogue is nominally magnitude limited to $b_J = 17.15$, in practice any galaxy catalogue is additionally limited by surface brightness. We plan to study surface brightness selection effects for APM scans using deep CCD frames and simulations; the results of this analysis will be published elsewhere. For now, note that the APM detection threshold corresponds to $\mu_J \approx 24.5$–25 b_J mag arcsec^{-2} (Maddox *et al.* 1990a), but the background subtraction employed by APM scans may reduce the effective surface brightness limit for extended images.

Seventy-one of our survey galaxies have already had redshifts measured as part of the Southern Sky Redshift Survey (da Costa *et al.* 1988). The sky distribution of all survey galaxies

is plotted in Figure 12.

4.   Observations

Observations were carried out over a total of 56 nights from November 1987 to November 1991 using the Dual-Beam Spectrograph (DBS) at the Nasmyth focus of the Australian National University (ANU) 2.3m telescope at Siding Spring. The spectrograph and detector systems have been described fully by Rodgers *et al.* (1988a,b), but the salient features are briefly mentioned here. A dichroic beam splitter behind the spectrograph slit divides the light into red and blue beams at about 5800Å. This enables the optics in each arm of the spectrograph to be optimized for their respective wavelength ranges, making this an extremely efficient instrument. The blue and red gratings are independently changeable. We used 600 l/mm gratings for both, giving a wavelength range of 3700–5000 Å in the blue and 6300–7600 Å in the red, both with dispersion $\approx$ 1Å/pixel. For galaxies at low redshift, the blue range includes the prominent H and K absorption features, the 4000Å break and O_{II} emission at 3727Å. The red range includes $H\alpha$, N_{II} and S_{II} emission lines. The two detectors are large-format photon-counting arrays (LFPCAs), consisting of micro-channel plate intensifier tubes, each optimized for the expected wavelength range, and each fibre-optically coupled to two Fairchild CCD chips. Since the fibre-bundle pairs do not butt up perfectly, (Rodgers *et al.*, 1988b), there is a small gap in wavelength coverage of about 4360–4370 Å in the blue, and 7000–7020 Å in the red. The CCDs are clocked out at a frame rate of 12ms into an incrementing memory, which can be read out whilst an exposure is in progress.

We used a slit width of $\sim$ 8 arcsec, leading to a spectral resolution, measured from the full-width at half-maximum of the arc lines, of about 5Å. Most galaxy exposures were 600s, but the exposure was stopped before this if strong emission lines appeared sooner. Up to seventy galaxies were observed each clear night. An automated field rotator was working on the spectrograph from the 1988 observing season, enabling us to occasionally rotate the slit in order to observe an additional non-survey galaxy at the same time as a survey object. These additional objects are not included in the current catalogue. Standard radial velocity stars and bright galaxies (taken from the *Astronomical Almanac*) were observed during twilight to be used as template spectra for cross-correlation velocity calculation. Helium-argon and neon arc spectra were taken about once every three galaxy exposures to tie down any drifts in the wavelength scale due to flexure of the spectrograph. Flat-field exposures using a quartz-iodine lamp were left running during the day.

5.   Reduction of Images to Calibrated Spectra

The main steps in the reduction of raw image data to calibrated spectra are (i) extracting a subsection from each image, (ii) flat-fielding, (iii) correcting image distortion, (iv) sky subtraction and (v) wavelength calibration. This preliminary reduction was carried out with the Figaro

data reduction package. Due to the slight gaps in wavelength coverage mentioned above, each of the four PCA images was reduced separately to yield four spectra for each observation. The flat-fielding and distortion correction were performed individually for each nights observing.

The central 200 rows from each PCA image were extracted. This area is large enough to include a generous amount of sky, yet excludes the field edges where vignetting of the flat-field lamp is considerable. The data in these extracted frames has wavelength increasing in the x direction (from left to right), the y direction maps position along the slit.

5.1. Flat-Fielding

A flat-field frame was made for each night by co-adding the quartz-iodine flat-field exposures left running during the daytime either side of the night in question. Each co-added flat-field pixel has several thousand counts and so Poisson errors in photon counting statistics cause only about 1% variation. The flat-fields show large-scale variations: in the x (wavelength) direction due to the spectral continuum of the halogen lamp and in the y (spatial) direction due to vignetting of the flat-field lamp towards the ends of the slit. This vignetting is not seen in sky images. In order to remove these gradients, a 15×15 pixel median filter was applied to the flat-field images. This yields a smoothed flat-field, containing only large-scale gradients. Dividing this into the original flat-field removes these gradients, leaving a final flat-field frame with just small scale variations in pixel intensity. Each object and arc frame is divided pixel-for-pixel by this flat-field.

5.2. Distortion Correction

Since the 2-d images are collapsed in the y direction to yield spectra, it is first desirable to make lines of constant wavelength (*e.g.* arc and sky emission lines) as vertical and straight as possible. The usual way of doing this is to make a fit to the arc lines for each row of the image, and then perform a 2-d wavelength 'scrunch'. This was tried, but it soon became apparent that the signal-to-noise for our arc lines was too low to provide a good fit to each row. We also tried a 2-d scrunch on summed groups of 10 rows. While this improved the fit to each group of rows, discontinuities became visible between them—there is no continuity requirement in the Figaro 2-d arc fitting program.

Therefore we used the Figaro routines SDIST and CDIST (originally intended for correcting 'S-Distortion' in RGO Spectrograph images) to take out the distortion. To obtain good signal to noise, the arc frames for each nights observing were co-added. The resultant image was rotated through 90°(so that the lines were horizontal) and displayed on an interactive image display. The cursor was used to select up to ten lines. Using these cursor positions, SDIST attempts to follow each line and fits a 10th-order polynomial. This particular choice of polynomial order is not important, but a high order fit is needed in order to follow occasional sharp kinks in the arc

lines due to shear faults in the fibre-optic couplers. The coefficients for each line are output to a file. Any lines which could not be followed successfully, eg. because they were too weak, were deleted. CDIST was then used to apply the distortion correction to each arc and object frame taken that night by fitting a 3rd-order polynomial perpendicular to the followed lines—a sort of '1.5-d' smoothing. Since smooth curves (polynomials) are fitted to the arc lines, no discontinuities are introduced.

5.3. Sky Subtraction

Each object exposure PCA image was displayed on an image display and a cursor used to delimit those rows containing light from the object. Two sky regions, one above and one below the object, were also identified, being careful to ensure that they were not contaminated by any other object that might have entered the slit during the exposure. These regions were collapsed in y and normalised by the number of rows to yield 1-d spectra. The mean of the two sky spectra was subtracted from the object spectrum to finally yield a sky-subtracted spectrum. The summed sky spectra were also saved so that they could be used to test the wavelength scale zero-points (§6.3.1).

5.4. Wavelength Calibration

The flat-fielded, distortion corrected arc frames were collapsed to spectra. For the first arc frame taken each night, the arc lines were identified interactively with the Figaro ARC program, with reference to the RGO CuAr wavelength identification guide (Schinckel, Phillips and Hill 1982). Between 5 and 14 lines were identified per CCD frame and a second or third-order polynomial fitted to obtain a pixel to wavelength calibration. The rms error of the fit ranged from 0.06 to 0.12 Å.

Once one arc spectrum had been calibrated in this way, the fit was used to perform an automatic fit to other arcs taken the same night. The rms wavelength error for each line (summed over arcs) and for each arc (summed over lines) was calculated to detect any deviations from a good fit. Any drift in wavelength scale during the night was monitored by calculating the fitted wavelength at three fixed pixel positions on each frame. On some nights we see a systematic drift in the blue wavelength scale from the beginning to the end of the night amounting to as much as 200 km/s. It thus seems likely that there may be some slippage in the blue grating angle. Between adjacent arc exposures (taken at roughly half-hourly intervals) the drift is less than 40 km/s, or occasionally larger when the spectrograph was rotated. Flexure in the red arm of the spectrograph is generally less than 10 km/s between adjacent arcs. In order to minimize the effects of spectrograph instability, the wavelength scale for each object spectrum was obtained by interpolating between the two arc fits taken either side, weighted inversely by the distance in

exposure sequence number from the object exposure. A further check on wavelength calibration, using sky lines, is described in §6.3.1.

6. Redshift Measurement

Redshifts were measured both by fitting individual emission lines and by cross-correlation with template spectra to determine redshifts from absorption features.

6.1. Emission-Line Velocities

Any obvious emission-line galaxies were noted down whilst making the observations, but in order to get as many emission-line velocities as possible from our sample, *all* galaxy spectra were checked for emission-lines. Each spectrum was displayed and any prominent emission-line was indicated with the cursor, and its rest wavelength typed in. The line peak was fitted by convolution with the derivative of a Gaussian; the same algorithm that is used by the Figaro arc-line fitting program. Once one line had been identified in this way, its redshift was used to predict the location of other emission lines from a table of rest wavelengths (Table 1). If a peak was found near the expected position it was marked and the user given the choice of whether to accept the line, or reject it as a noise peak. In this way, up to 12 lines were identified per emission-line galaxy spectrum. Out of 1747 galaxy spectra we observed, 1079 showed two or more emission lines. An rms velocity error for each galaxy was calculated from the variance between the redshifts of each line. The median error was 20 km/s.

Figure 1 shows our 'best' emission-line galaxy which shows at least twelve emission lines, ten of which are identified. The numbers labeling the lines are the peak wavelength (in Å) followed by the redshift velocity $c\Delta\lambda/\lambda$.

6.2. Cross-Correlation Velocities

We used the cross-correlation method of Tonry and Davis (1979) to determine redshifts of galaxies from their absorption features. The power of this method lies in the fact that all of the spectral features (after filtering out high frequency noise and the low frequency continuum) are used, not just selected lines. It also provides a reliable estimate of velocity error in most cases.

Cross-correlation methods were tried for the blue and red spectra separately. Since there is a large discontinuity of about 1300 Å between the blue and red spectra, there is little advantage in joining the two together—it would take a redshift $z \sim 0.3$ to shift a feature from the end of the blue to the start of the red wavelength range. It was found that the red range was too highly contaminated by sky lines (Fig. 9) to provide a reliable cross-correlation velocity, and so only the

blue spectra were used to obtain cross-correlation velocities.

6.2.1. Theory

Consider a galaxy spectrum $g(n)$ and a template spectrum $t(n)$, with Fourier transforms $G(k)$ and $T(k)$ respectively, both binned into N bins on a logarithmic wavelength scale

$$n = A \ln \lambda + B. \tag{3}$$

A redshift $z = d\lambda/\lambda$ would thus cause a shift in bins

$$\delta = A \ln \left(\frac{\lambda + d\lambda}{\lambda} \right) = A \ln(1 + z), \tag{4}$$

and so

$$z = e^{\delta/A} - 1. \tag{5}$$

Thus while a given redshift will shift the logarithmically binned spectra uniformly, the shift δ is *not* a linear function of z. The cross-correlation $c(n)$ of these spectra is defined by

$$c(n) = \frac{1}{N \sigma_g \sigma_t} \sum_i g(i) t(i - n), \tag{6}$$

or in Fourier space

$$C(k) = \frac{1}{N \sigma_g \sigma_t} G(k) T^*(k), \tag{7}$$

where σ_g is the rms of $g(n)$, σ_t the rms of $t(n)$ and $*$ denotes complex conjugation.

If $g(n)$ were identical to $t(n)$ but shifted by δ bins, then $c(n)$ would would have a peak of unit height shifted by δ from the origin. In practice, mis-match between the two spectra will give rise to spurious, noise peaks in the cross-correlation which can distort the true peak and so lead to error in the measurement of δ. Tonry and Davis (1979) define a parameter r

$$r = \frac{h}{\sqrt{2} \sigma_a} \tag{8}$$

which is the ratio of the height h of the true peak to the height $\sqrt{2} \sigma_a$ of an average, noise peak, where σ_a is the rms of the antisymmetric part of $c(n)$. Under reasonable assumptions[1], they show that the mean error ε in δ is given by

$$\varepsilon = \frac{N}{8B} \frac{1}{1 + r}, \tag{9}$$

---

[1](i) The measured cross-correlation is the sum of a perfect (symmetric) correlation of a shifted, broadened template against template plus an asymmetric remainder function.

(ii) The remainder function has symmetric and antisymmetric components which are uncorrelated and of equal *rms*.

(iii) The peaks in $c(n)$ may be represented by parabolas.

(iv) The true peak is only significantly distorted by the nearest spurious peak.

where B is the wave number at which $C(k)$ reaches its half-maximum. In practice, B is poorly determined and so we have determined the constant C in

$$\varepsilon = \frac{C}{1 + r} \tag{10}$$

by performing Monte Carlo experiments (see §6.2.4).

6.2.2. Spectrum Preparation

Firstly the two blue spectra of each observation were continuum subtracted by fitting and subtracting a fourth-order polynomial (with outlyer rejection to avoid the effects of strong emission or absorption features) and then subtracting the mean. Since any strong emission lines would dominate the cross-correlation, any spectrum peaks exceeding the continuum by 5 times the rms noise were removed by setting all points in the peak to the continuum level (now zero). Five per-cent of each end of each spectrum was apodized with a cosine bell so that the data went smoothly to zero. The two spectra were then adjoined and 'scrunched' onto a logarithmic wavelength scale from 3700–5000 Å with 2048 bins, giving the scrunched spectrum a velocity increment of about 44.1 km/s/bin. Remaining continuum variations and high frequency noise were removed by applying a middle-pass filter $f(k)$ to the spectrum in Fourier space,

$$f(k) = e^{-(k/u)} - e^{-(k/v)}. \tag{11}$$

The optimum filter parameters u and v were determined empirically by minimizing the difference between our cross-correlation velocities and published velocities (§6.2.3).

As an illustration of the stages in spectrum preparation Fig. 2 shows (a) the two blue wavelength-calibrated spectra for a survey galaxy; (b) the same spectra after they have been continuum-subtracted, cosine-belled and adjoined and (c) after rebinning onto a logarithmic wavelength scale and Fourier filtering.

6.2.3. Zero-Pointing the Templates

Tonry and Davis (1979) have pointed out that some care must be taken in adjusting the template spectra to zero redshift, as off-centre guiding of a stellar image smaller than the slit width can introduce systematic errors of up to 150 km/s for our resolution of 5Å. We therefore follow a procedure outlined by Tonry and Davis in which all of the observed template spectra which have accurate published velocities v_{pub_i} are put onto a common velocity system. Each spectrum is correlated with all of the others, thus generating a table of *relative* velocities r_{ij}. If each object is *assigned* a velocity v_i, then we can calculate another table of relative velocities whose entries are

$v_i - v_j$. We wish to minimize the expression

$$\chi^2 = \sum_{i,j} \frac{[r_{ij} - (v_i - v_j)]^2}{\varepsilon_{ij}^2}, \tag{12}$$

with respect to the v_i, subject to the constraint

$$\sum_i v_i = \sum_i v_{\mathrm{pub}_i}, \tag{13}$$

where ε_{ij} is the cross-correlation velocity error, and the constraint ensures that the sum of assigned velocities is equal to the sum of published velocities. We thus have

$$\frac{\partial \chi^2}{\partial v_k} = 2 \sum_{i \neq k} \frac{r_{ik} - (v_i - v_k)}{\varepsilon_{ik}^2} = 0, \tag{14}$$

and so

$$\sum_{i \neq k} \frac{v_i}{\varepsilon_{ik}^2} - v_k \sum_{i \neq k} \frac{1}{\varepsilon_{ik}^2} = \sum_{i \neq k} \frac{r_{ik}}{\varepsilon_{ik}^2} \tag{15}$$

for all k. In matrix notation $\mathbf{A}\mathbf{x} = \mathbf{b}$ this last expression can be rewritten

$$\begin{bmatrix} -\sum 1/\varepsilon_{i1}^2 & 1/\varepsilon_{21}^2 & \cdots & 1/\varepsilon_{N1}^2 \\ 1/\varepsilon_{12}^2 & -\sum 1/\varepsilon_{i2}^2 & \cdots & 1/\varepsilon_{N2}^2 \\ \vdots & \vdots & \vdots & \vdots \\ 1/\varepsilon_{1N}^2 & 1/\varepsilon_{2N}^2 & \cdots & -\sum 1/\varepsilon_{iN}^2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} = \begin{bmatrix} \sum r_{i1}/\varepsilon_{i1}^2 \\ \sum r_{i2}/\varepsilon_{i2}^2 \\ \vdots \\ \sum r_{iN}/\varepsilon_{iN}^2 \\ \sum v_{\mathrm{pub}_i} \end{bmatrix}, \tag{16}$$

which is a standard simultaneous equations problem. For a system of N templates, $\mathbf{A}$ is an $(N+1) \times N$ matrix, in which the $(N+1)^{\mathrm{th}}$ row enforces the constraint (13).

By considering all pairs of template cross-correlations, we can calculate an internal error σ_{int} from the cross-correlation minus best-fit velocities and an external error σ_{ext} from the cross-correlation minus published velocities. The filter parameters u, v in equation (11) were varied so as to minimize the external error σ_{ext}. Table 2 shows the internal and external errors obtained for a system of 36 templates for several values of u, v. The errors are clearly fairly insensitive to the exact filter parameters used, except that taking a value for v as high as 30 removes too much low-frequency signal. The optimal values appear to be around $(u, v) = (80, 20)$.

Since 36 templates is a rather excessive number to cross-correlate with over 1700 galaxy spectra, those templates with best-fit velocities differing by more than 50 km/s from the published velocities were rejected and the remaining 23 templates shifted to zero velocity. Repeated observations of the same template were co-added, giving 15 different template spectra. The templates used are listed in Table 3. A composite template spectrum was also obtained by co-adding these 15 spectra; this composite template is shown in Figure 3.

6.2.4. Calibration of Velocity Error

The constant C in equation (10), relating the probable velocity error ε to the cross-correlation r factor, was calibrated by taking a high signal-to-noise stellar template and adding Gaussian random noise to each spectrum pixel. This noise degradation was repeated to give 2000 degraded spectra. The original spectrum was then cross-correlated with each of the degraded spectra. Figure 4 plots the velocity shift Δv against r for each correlation. The solid line shows rms(Δv) measured for 10 equally spaced r-bins from $r = 2$ to $r = 12$. From this data we find $C = 210 \pm 20$ km/s; the dashed line shows the curve $\varepsilon = C/(1 + r)$, which is clearly a reasonable fit to the data.

6.2.5. Correlation of Galaxy Spectra with Templates

Each galaxy spectrum, prepared as described in §6.2.2, was correlated with each of the 16 template spectra, including the composite template, and the highest correlation peak identified in the range $-5,000$ to $+40,000$ km/s. Figures 5–7 show spectra and cross-correlation plots for representative good, average and poor signal-to-noise galaxies. Whilst the correct correlation peak is clear for the good and average quality spectra, it is not at all obvious in the poor quality case. In this case, choosing the wrong peak would give a totally spurious velocity, whose error would be grossly underestimated by equation 10. Therefore, for each galaxy, the redshifts from the *three* galaxy-template correlations giving the highest r values were output to a file, enabling an extra check to be made on redshift reliability. Denoting the best (highest r) velocity v_1, and the second and third best velocities v_2 and v_3, the median velocity shifts are $|v_1 - v_2| = 26$ km/s and $|v_1 - v_3| = 27$ km/s.

Figure 8 plots the rms velocity dispersion amongst the three best templates against r for each galaxy spectrum. Most velocity errors are less than 160 km/s, but about 18% have a dispersion more than 1000 km/s, with very few points in between. These discrepant points have a range in r up to about 4.0 while some correlation velocities are in agreement for r as low as 2.5. This illustrates a serious problem with the cross-correlation technique for determining galaxy redshifts—a spurious correlation peak can be chosen whilst still giving a reasonable r for some spectra. By comparing the velocities obtained with three different templates, we can detect at least some instances where a spurious peak has been found. It is possible that a spurious feature in a galaxy spectrum may give the same false correlation peak with all three templates. However, noise in our PCA spectra is dominated by random photon-counting errors, rather than systematic errors as in CCD images, and so a significant spurious feature (cf. Fig. 3) is unlikely. In §7.1 below we compare velocities determined from cross-correlation and from emission lines to show that finding the same three false peaks is indeed very unlikely.

6.3. Corrections to Redshifts

6.3.1. Correlation of Sky Spectra

As a check on wavelength calibration, the sky spectra saved from each exposure were correlated with a sky template spectrum in order to test for any spurious velocity shifts. The sky spectra were prepared in the same way as the galaxy spectra (§6.2.2), except that the sky emission lines were not removed. In order to form a template spectrum, a sky spectrum from each of the first twenty clear nights was taken. These 20 sky spectra were cross-correlated and shifted to zero velocity as for the stellar and galaxy templates (§6.2.3). They were then added to give an extremely high signal-to-noise template sky spectrum. This process was repeated for the red spectra; the sky templates are plotted in Figure 9. It is clear from this figure why the red galaxy spectra are unsuitable for finding cross-correlation velocities—the red sky lines are about twenty times as bright as those in the blue. Even if the sky-line regions of the spectrum were interpolated over, there would be very little spectrum left for cross-correlation.

Each blue and red sky spectrum was correlated with the blue or red sky template. Figure 10 is a scatter-plot of velocity shift against r for (a) the blue and (b) the red sky spectra. The median velocity shifts are -0.2 and -0.5 km/s respectively, confirming that the templates were accurately zero-pointed. The blue correlation velocities show an rms of 41 km/s and are uniformly distributed about zero but the red correlations show streaks of roughly constant Δv, separated by up to 150 km/s. This is probably a consequence of there only being five arc lines on each red frame leading to a slightly unstable wavelength calibration. It can be seen from the r values that nearly all sky spectra give an excellent correlation with the template. The few that give poor correlations are the sky spectra for bright standard stars, where the combination of a neutral-density filter and short exposure time gives a noisy sky spectrum.

The cross-correlation galaxy velocities were corrected by subtracting the blue sky velocity shift for that exposure. Emission-line velocities were corrected by subtracting the blue sky velocity shift for lines measured in the blue ($\lambda < 5000$ Å), and subtracting the red sky velocity shift for lines measured in the red ($\lambda > 6400$ Å).

6.3.2. Heliocentric Correction

The final correction applied to the measured redshifts was to convert from a geocentric to a heliocentric frame. This correction was calculated using the subroutine SLA_ECOR in the Starlink SLALIB library, given the date and time of observation, and amounts to a correction of up to ± 30 km/s.

7.   Redshift Statistics, Accuracy and Reliability

7.1.   Comparison of Emission-line and Absorption Velocities

Before comparing the emission-line ($v_{\rm em}$) and absorption ($v_{\rm abs}$) velocities, we check that the absorption velocities for the three best templates agree to within 160 km/s (§6.2.5). If they do, and we have an emission-line velocity for that galaxy, then $v_{\rm abs}$ and $v_{\rm em}$ always agree to within 500 km/s.

Emission-line velocities can differ by up to ~ 500 km/s from absorption-feature velocities, since much of the energy in emission-lines comes from $H_{\rm II}$ regions which maybe orbiting the nucleus of a spiral galaxy. If the galaxy is edge on, then this can lead to the difference in absorption and emission velocities. Therefore, in cases where the three template velocities are in disagreement, if we measure an emission-line velocity and the closest absorption velocity is within 500 km/s, then we assume that that absorption velocity is correct. If the nearest absorption velocity is more than 500 km/s from the emission-line velocity, we take $v_{\rm em}$ to be the more accurate. If there is no emission-line velocity, then we take the best (highest r) absorption velocity, and flag it as uncertain. Only 2 galaxy redshifts fall into this category.

In Figure 11 we plot the frequency distribution of (a) $v_{\rm abs}$, (b) error in $v_{\rm abs}$ (from equation 10), (c) $v_{\rm em}$, (d) error in $v_{\rm em}$ (rms between lines) and (e) $v_{\rm abs} - v_{\rm em}$. Figure 11(f) is a scatter-plot of $v_{\rm abs} - v_{\rm em}$ plotted against $v_{\rm abs}$. We have excluded absorption velocities flagged as uncertain from these plots.

Table 4 shows the numbers of redshift observations falling into different categories according to presence or absence of emission-line and reliable absorption velocities. More than 99% of galaxies have a reliable absorption or emission-line velocity, and 47% have both. Cross-correlation plots for the seven objects with no reliable velocity were checked by eye. For five, no clear peak could be discerned. For the remaining two, a likely peak could be identified, but the redshift so obtained was flagged as unreliable.

Table 5 compares the velocities obtained from absorption and emission-line features. Whilst the median velocity for emission-line galaxies is about 1500 km/s less than for absorption-feature velocities, the median difference for galaxies with absorption *and* emission-line velocities of -20 km/s is negligible compared with the rms difference of 124 km/s.

7.2.   Repeated Observations

During the course of observations, 134 galaxies were observed more than once (for example if one observation was made under poor weather conditions). By comparing the redshifts measured with repeat observations, we can obtain an estimate of redshift reliability and accuracy. This will be a pessimistic estimate, since only poor quality observations were repeated. Had high quality

observations also been repeated, we would expect much smaller errors.

The median dispersion between absorption-velocity estimates is 1190 km/s. However, 80 of the repeated observations have absorption velocities flagged as unreliable (the best 3 template correlation velocities differ by more than 160 km/s). If these are excluded, then the remaining 54 repeated observations have a median dispersion of 55 km/s, consistent with the errors estimated from equation (10). The median emission-velocity dispersion is 25 km/s.

Twenty-one galaxies had redshifts measured by us *and* the Southern Sky Redshift Survey (SSRS). Excluding one discrepant pair[2] ($v_{SSRS} = 2669$, $v_{APM} = 4$ km/s), we find $v_{SSRS} - v_{APM} = 9 \pm 115$ km/s. The estimated error for SSRS velocities is 40 km/s.

8.   Catalogue Summary

We have described the construction of a new sparse-sampled redshift survey of 1797 galaxies brighter than $b_J = 17.15$ covering 4300 square degrees of the southern sky. Figure 12 shows the redshift survey galaxies in an equal-area projection centred on the south galactic pole.

Out of our total sample of 1797 galaxies, we attempted to measure redshifts for 1747 using both emission lines and absorption features. All but 7 of these yielded a reliable redshift. Redshifts for the remaining galaxies were obtained from the literature: fifty from the Southern Sky Redshift Survey (da Costa *et al.* 1988, SSRS), one from Dressler and Shectman (1988) and three from Huchra's ZCAT redshift compilation. Our survey is thus 99.6% complete. Sky spectra were used to zero-point the velocity for every observation. Reliability of the cross-correlation velocities is ensured by comparing velocities from three different templates. All emission line redshifts are based on at least two lines.

Figure 13 plots the galaxy distribution in right-ascension–redshift space for four declination ranges. Table 6 presents a small sample of the data in the Stromlo-APM Redshift Survey. The complete catalogue will be available from the Astronomical Data Centre (http://adc.gsfc.nasa.gov/). Each column in the table is described below.

(1) **Name:** Galaxy naming follows the same convention as the APM Bright Galaxy Catalogue (Loveday 1996). Each galaxy name is composed of the survey field number and the x, y position of the galaxy on the plate—this should ease location of any particular galaxy on the plate material. The first 3 digits are the SERC field number. The second set of digits are the x-position in millimetres from the centre of the plate (actually the APM scan centre). These are preceded by a '+' sign for galaxies to the right (west) of the plate centre or by a '−' sign for galaxies to the left (east) of centre. The final 3 digits are the y position, again

---

[2]This object was inspected on the plate and it was noticed that there was a star close to the galaxy nucleus. Presumably we measured the redshift of the star.

in mm from the plate centre. A preceding '−' indicates galaxies above (north) of the plate centre, '+' indicates galaxies below (south) of the centre.

(2), (3) RA, dec: Right ascension (hours, minutes, seconds) and declination (degrees, arcminutes, arcseconds) in 1950 coordinates.

(4) b_J**:** Matched, saturation corrected b_J magnitude. Note that although two decimal places of precision are quoted, comparison with CCD photometry (Paper 1) shows that individual galaxy magnitudes are accurate only to ≈ 0.3 mag. Note also that the magnitudes shown here differ by around 0.05 mag from the same galaxies in APM-BGC. This is because the Stromlo-APM survey includes five extra Schmidt fields not included in APM-BGC, and the plate zero-points change slightly when these five extra fields are included in the matching solution.

(5), (6) maj, min: Major and minor diameter in arcseconds at threshold isophote.

(7) p.a.: Position angle in degrees measured clockwise from south-north line.

(8) Cl: Galaxy classification code: 1 = E, 2 = S0, 3 = Sp, 4 = Irr/Pec, 5 = unsure. Those galaxies that were only found by cross-checking with the ESO survey (Lauberts 1982) have had 10 added to the classification code. Those galaxies classified by SJM have had 20 added; all others were classified by JL. Note that the distinction between elliptical and lenticular galaxies is rather poor (Loveday 1996), and it is suggested that classes 1 and 2 are combined in any analysis that depends on morphological type.

(9) cz**:** Heliocentric recession velocity in km/s.

(10) err: Estimated error in heliocentric recession velocity in km/s.

(11) src: Source of redshift: −1 = unreliable, 0 = not available, 1 = absorption features, 2 = emission lines, 3 = absorption plus emission features, 4 = SSRS, 5 = Dressler and Shectman 1988, 6 = Huchra's ZCAT redshift compilation.

Tables

Table 1: Galaxy emission-line rest wavelengths in Å (Sandage 1975).

Rest Wavelength	Line
3727.3	[O$_{\rm II}$]
3868.74	[Ne$_{\rm III}$]
4101.74	H$_\delta$
4340.47	H$_\gamma$
4861.33	H$_\beta$
4958.91	[O$_{\rm III}$]
5006.84	[O$_{\rm III}$]
5875.63	He$_{\rm I}$
6548.10	[N$_{\rm II}$]
6562.82	H$_\alpha$
6583.60	[N$_{\rm II}$]
6717.00	[S$_{\rm II}$]
6731.30	[S$_{\rm II}$]

Table 2: Effects of varying filter parameters u, v on template correlation errors.

v	u	σ_{ext} km/s	σ_{int} km/s
20	150	98	40
30	100	2251	2203
20	100	97	36
10	100	154	122
20	**80**	**97**	**34**
15	80	98	33
20	60	98	33
20	50	99	33

Table 3: Stellar and galaxy templates used with spectral/morphological type and heliocentric redshift velocity (km/s).

Star	Type	v	Galaxy	Type	v
HD 4128	K1 III	+13.1	N1316	LAB(s)0p	+1774
HD 22484	F8 V	+27.9	N1380	LA0	+1809
HD 36079	G5 III	−13.5	N4697	E6	+1308
HD 150798	K2 IIb–IIIa	−3.7	N5102	LA0$^-$	+454
HD 168454	K2 III	−20.0	N5236	SAB(s)c	+518
HD 203638	K0 III	+21.9	N5643	SAB(rs)c	+1142
HD 213415	G5	+33.2	N5921	SB(r)bc	+1475
			I5267	LA(rs)0:	+1715

Table 4: Redshift observation statistics.

	N	%
Total	1747	
Reliable absorption velocities	1489	85.2
Emission-line velocity	1079	61.8
Reliable absorption velocity plus emission	825	47.2
Unreliable absorption velocity but with emission	254	14.5
Neither reliable absorption velocity nor emission	7	0.4

Table 5: Comparison of absorption and emission-line velocities (km/s).

Median v_{abs}	15526
Median v_{abs} error	38
Median v_{em}	14060
Median v_{em} error	20
Median $(v_{abs} - v_{em})$	-20
Mean $(v_{abs} - v_{em})$	-19
Rms $(v_{abs} - v_{em})$	124

Table 6: Subsample of the Stromlo-APM Redshift Survey Catalogue

Name	RA			Dec			b_J	maj	min	p.a.	Cl	cz	err	src
053+108-076	2	13	55.83	-68	34	5.4	16.26	33	26	127	2	18867	29	1
053+079-015	2	19	1.49	-69	44	33.7	16.96	32	14	86	22	22117	33	1
053+115-091	2	12	57.49	-68	16	23.6	16.57	37	18	104	23	19171	31	1
053-039+018	2	44	49.07	-70	23	58.1	17.12	32	10	20	23	18814	41	3
075+069-077	21	26	16.37	-68	39	53.3	16.32	32	19	59	4	11034	57	1
075-093+008	22	0	57.38	-70	13	10.7	15.93	81	11	144	3	8487	41	3
075-048-139	21	49	54.27	-67	31	26.6	16.84	25	20	145	21	20834	38	1
075-058-010	21	52	58.03	-69	55	41.9	15.05	60	36	47	2	8513	44	3
075+115+044	21	14	7.09	-70	49	58.6	15.90	42	21	58	3	13640	48	3
075-060+104	21	54	51.14	-72	2	23.1	16.03	35	28	47	3	20756	44	1
075+062-052	21	27	22.70	-69	7	59.1	16.67	42	10	87	23	12522	21	1
075-021+022	21	44	59.84	-70	33	6.3	16.80	38	13	32	23	10728	40	3
075-005-014	21	41	30.13	-69	52	25.6	16.83	28	19	85	23	17002	59	3
076+022-124	22	27	41.34	-67	47	50.5	16.22	31	21	11	3	12389	43	3
076-060-049	22	44	39.41	-69	10	1.2	14.33	64	45	177	3	3821	29	4
076-094-048	22	51	40.58	-69	7	59.3	17.14	25	15	64	23	22077	46	1
076-113-015	22	56	20.64	-69	43	6.6	16.48	41	17	52	23	3813	47	3
076+041-040	22	23	22.55	-69	20	11.6	16.65	28	24	62	21	20212	29	1
076+044+039	22	22	3.74	-70	49	19.5	16.45	33	20	161	23	19992	31	3
076-085+121	22	52	54.87	-72	17	6.8	17.09	32	13	92	21	22025	39	1

REFERENCES

Couch, W.J. and Newell, E.B., 1980, PASP, 92, 746

da Costa, L.N., Pellegrini, P.S., Sargent, W.L.W., Tonry, J., Davis, M., Meiksin, A. and Latham, D.W., 1988, ApJ, 327, 544 (SSRS)

Davis, M. and Huchra, J., 1982, ApJ, 254, 437

Davis, M. and Peebles, P.J.E., 1983, ApJ, 267, 465

Davis, M., Efstathiou, G., Frenk, C.S. and White, S.D.M., 1985, ApJ, 292, 371

de Lapparent, V., Geller, M.J. and Huchra, J.P., 1986, ApJ, 302, L1

de Lapparent, V., Geller, M.J. and Huchra, J.P., 1988, ApJ, 332, 44

Dressler, A. and Shectman, S.A., 1988, AJ, 95, 284

Groth, E.J. and Peebles, P.J.E., 1977, ApJ, 217, 385

Hamilton, A.J.S., 1993, ApJ, 417, 19

Huchra, J., Davis, M., Latham, D. and Tonry, J., 1983, ApJS, 52, 89

Kaiser, N., 1986, MNRAS, 219, 785

Lauberts, A., 1982, The ESO/Uppsala Survey of the ESO(B) Atlas, European Southern Observatory

Loveday, J., 1996, MNRAS, 278, 1025

Loveday, J., Efstathiou, G., Peterson, B.A. and Maddox, S.J., 1992a, ApJ, 400, L43

Loveday, J., Peterson, B.A., Efstathiou, G. and Maddox, S.J., 1992b, ApJ, 390, 338 (Paper 1)

Loveday, J., Maddox, S.J., Efstathiou, G., and Peterson, B.A., 1995, ApJ, 442, 457

Loveday, J., Efstathiou, G., Maddox, S.J., and Peterson, B.A., 1996, ApJ in press

Maddox, S.J., Sutherland, W.J. Efstathiou, G., and Loveday, J., 1990a, MNRAS, 243, 692

Maddox, S.J., Efstathiou, G. and Sutherland, W.J., 1990b, MNRAS, 246, 433

Maddox, S.J., Efstathiou, G., Sutherland, W.J. and Loveday, J., 1990c, MNRAS, 242, 43P

Marzke, R.O., Huchra, J.P., and Geller, M.J., 1994, ApJ, 428, 43

Peebles, P.J.E., 1973, ApJ, 185, 413

Peebles, P.J.E., 1980, The Large-Scale Structure of the Universe, Princeton University Press

Rodgers, A.W., Conroy, P. and Bloxham, G., 1988a, PASP, 100, 626

Rodgers, A.W., Van Harmelen, J., King, D., Conroy, P. and Harding, P., 1988b, PASP, 100, 841

Sandage, A., 1975, in Galaxies and the Universe, ed. Sandage, Sandage and Kristian, p 761

Schinckel, A.E., Phillips, M.M. and Hill, P.W., 1982, A Guide to Wavelength Identifications For the CuAr Lamp of the RGO Spectrograph, AAO

Seldner, M., Siebers, B., Groth, E.J. and Peebles, P.J.E., 1977, AJ, 82, 249

Shane, C.D. and Wirtanen, C.A., 1967, Pub. Lick. Obs., 22, Part 1

Tonry, J. and Davis, M., 1979, AJ, 84, 1511

White, S.D.M., Frenk, C.S., Davis, M. and Efstathiou, G., 1987, ApJ, 313, 505

Figure Captions

Fig. 1.— Calibrated spectrum of galaxy 251−009+023 showing at least twelve emission lines, ten of which are identified.

Fig. 2.— Illustration of the steps in spectrum preparation: (a) raw spectra, (b) after continuum subtraction and cosine-belling, and (c) after Fourier smoothing and 'scrunching'.

Fig. 3.— Composite template spectrum.

Fig. 4.— Noise degradation experiment to determine Δv as a function of r. The continuous line shows $+rms(\Delta v)$ and the dashed line shows the curve $\varepsilon = C/(1 + r)$.

Fig. 5.— Spectrum and cross-correlation plot for a good signal-to-noise galaxy.

Fig. 6.— Spectrum and cross-correlation plot for an average signal-to-noise galaxy.

Fig. 7.— Spectrum and cross-correlation plot for a poor signal-to-noise galaxy.

Fig. 8.— Scatter-plot of rms velocity dispersion amongst the three best galaxy-template correlation redshifts against r.

Fig. 9.— Template blue and red sky spectra after continuum subtraction, cosine belling, 'scrunching' and Fourier filtering.

Fig. 10.— Velocity shift against r for (a) blue and (b) red sky spectra.

Fig. 11.— Comparison of emission-line and absorption velocities.

Fig. 12.— Distribution of redshift survey galaxies in an equal-area projection centred on the SGP.

Fig. 13.— Distribution of survey galaxies in right-ascension–redshift space for four declination slices.
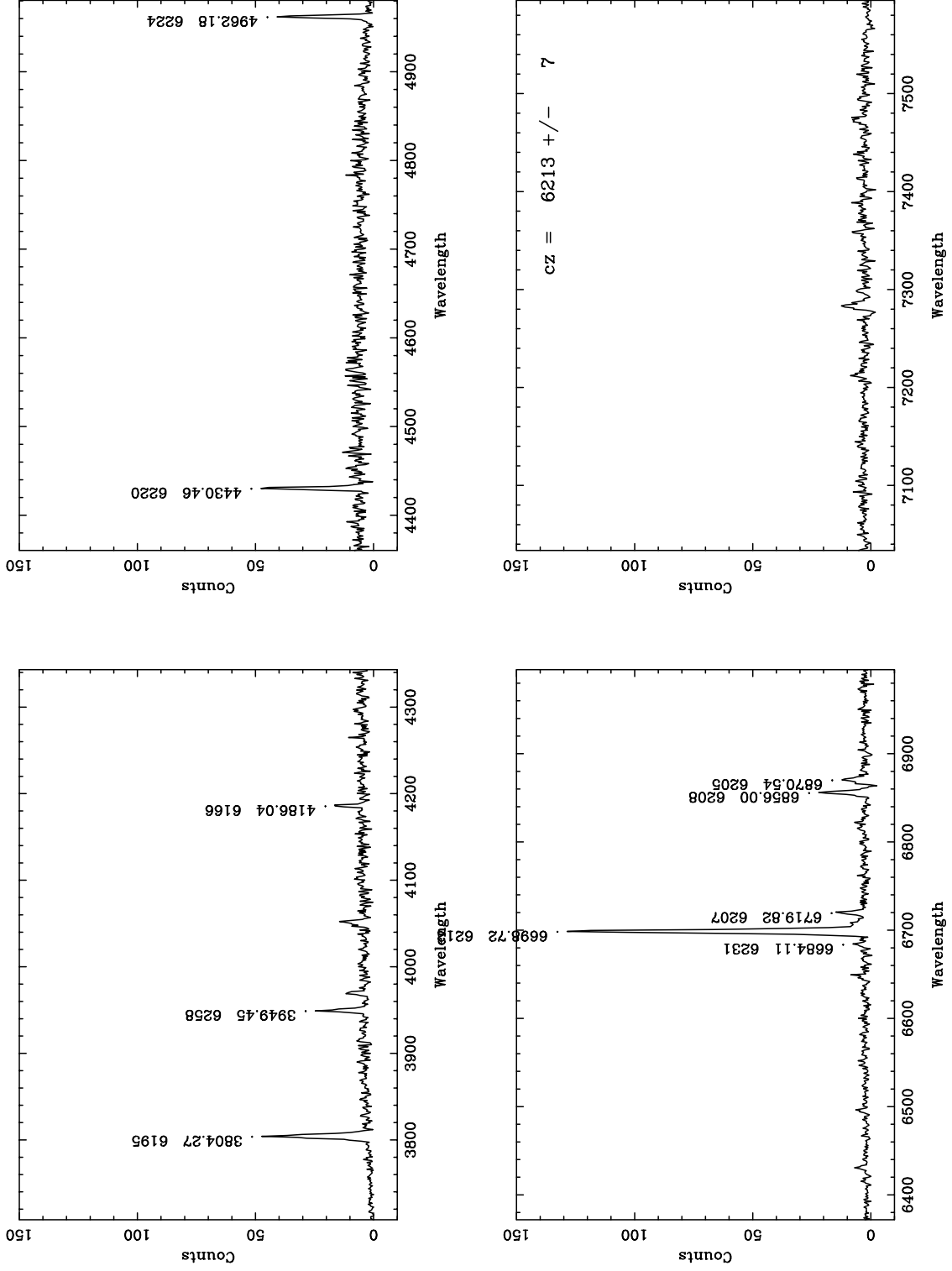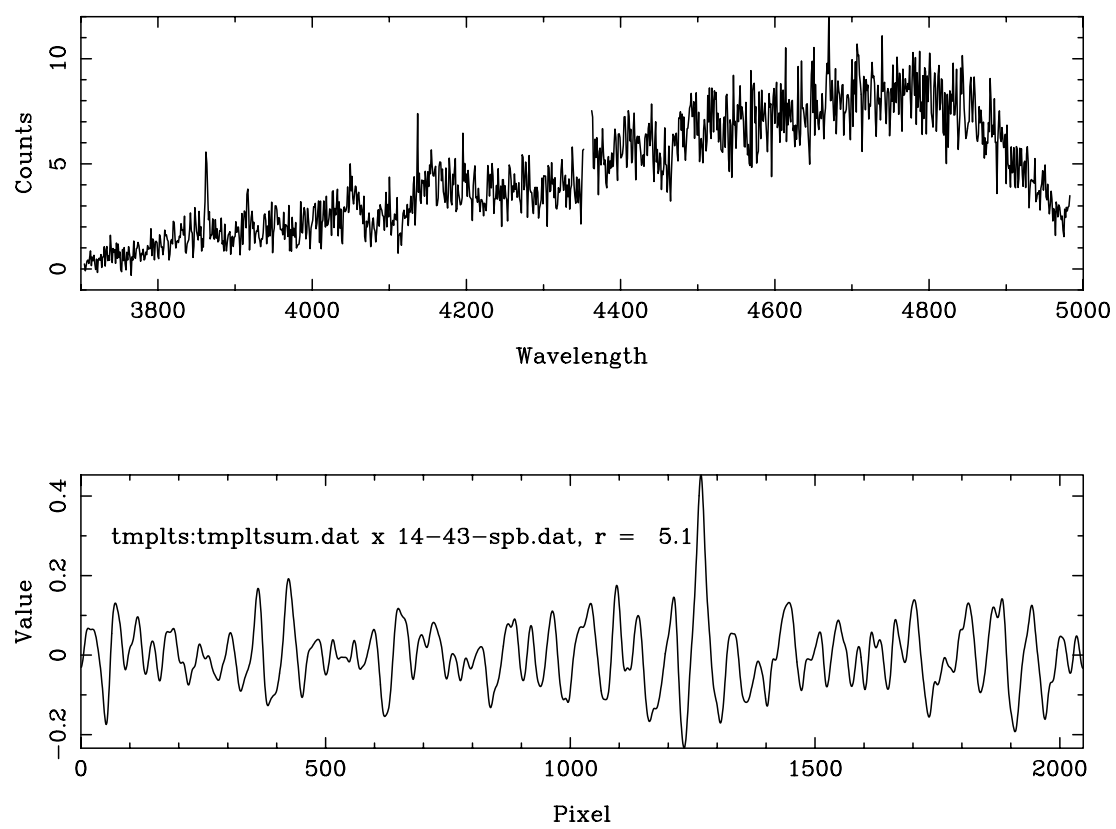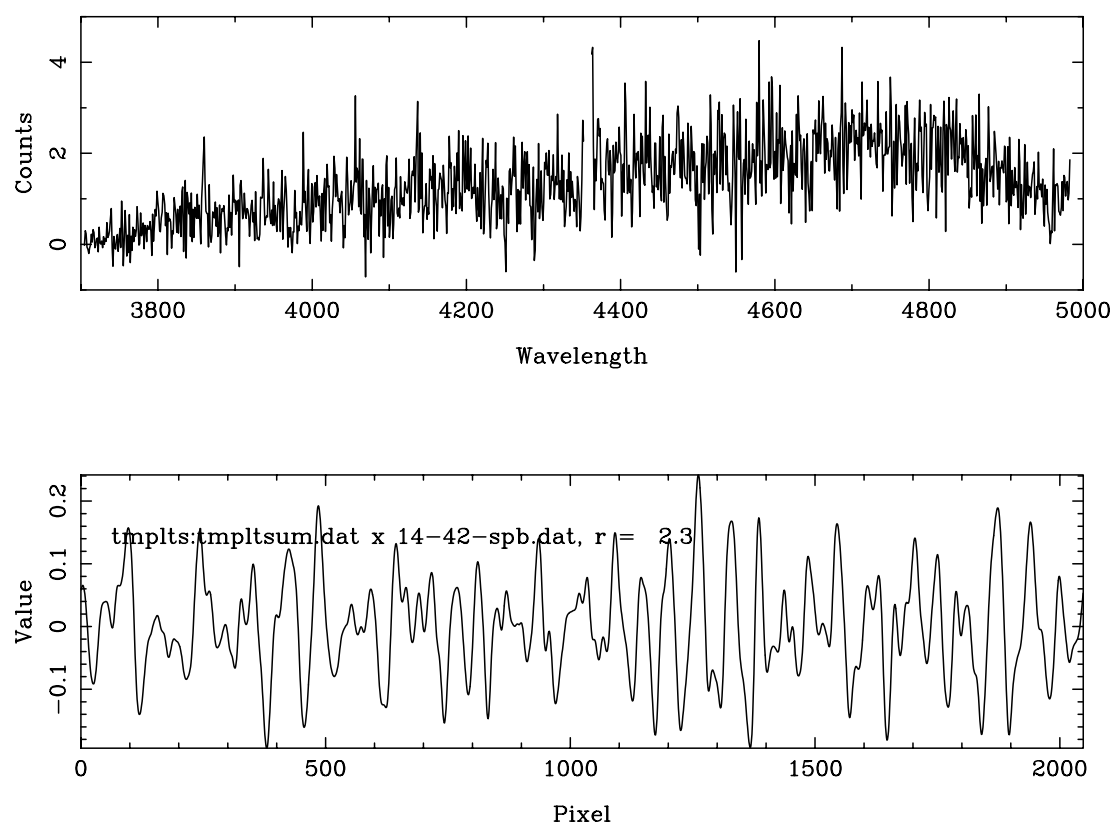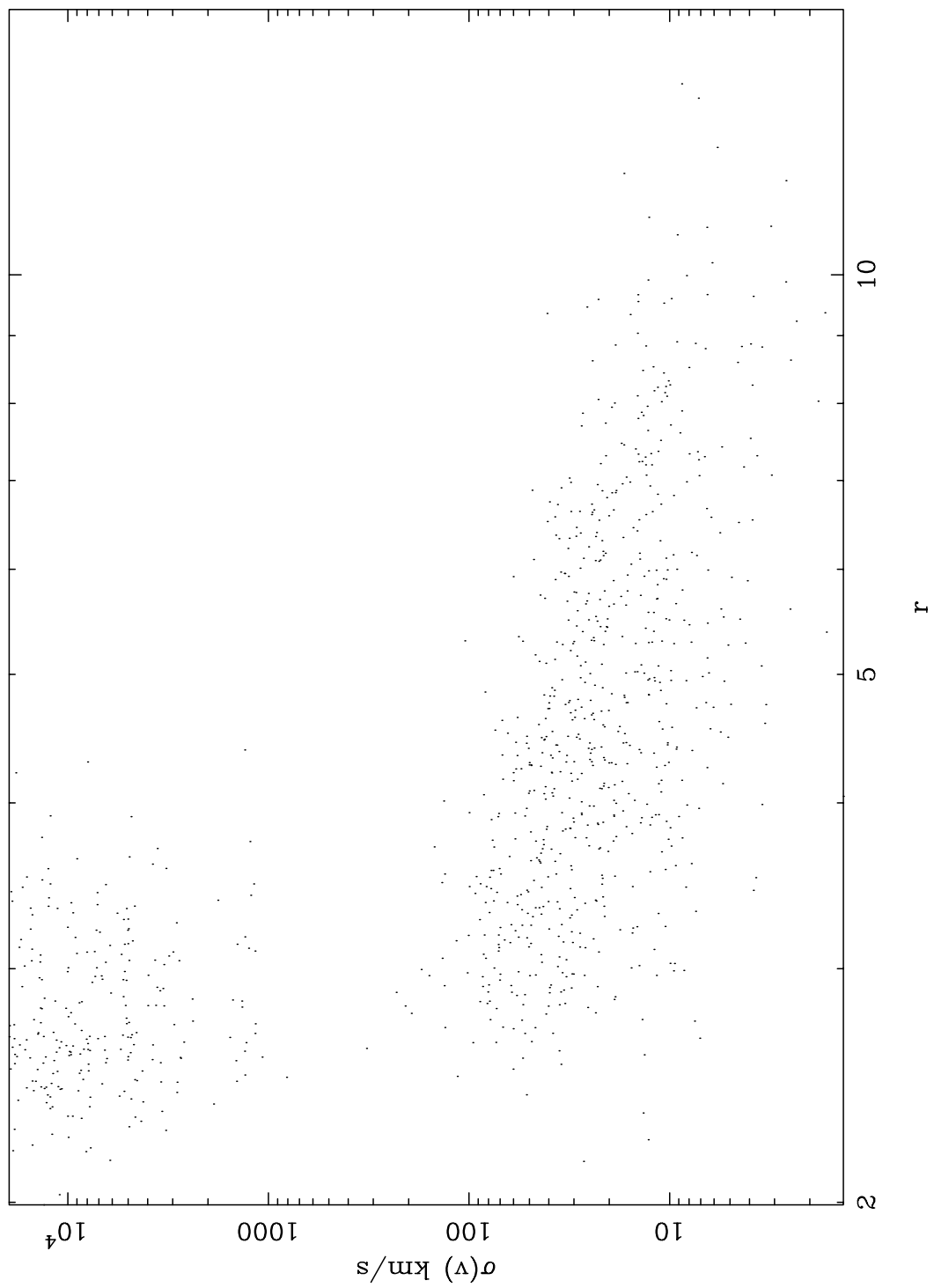
Fig. 1.—
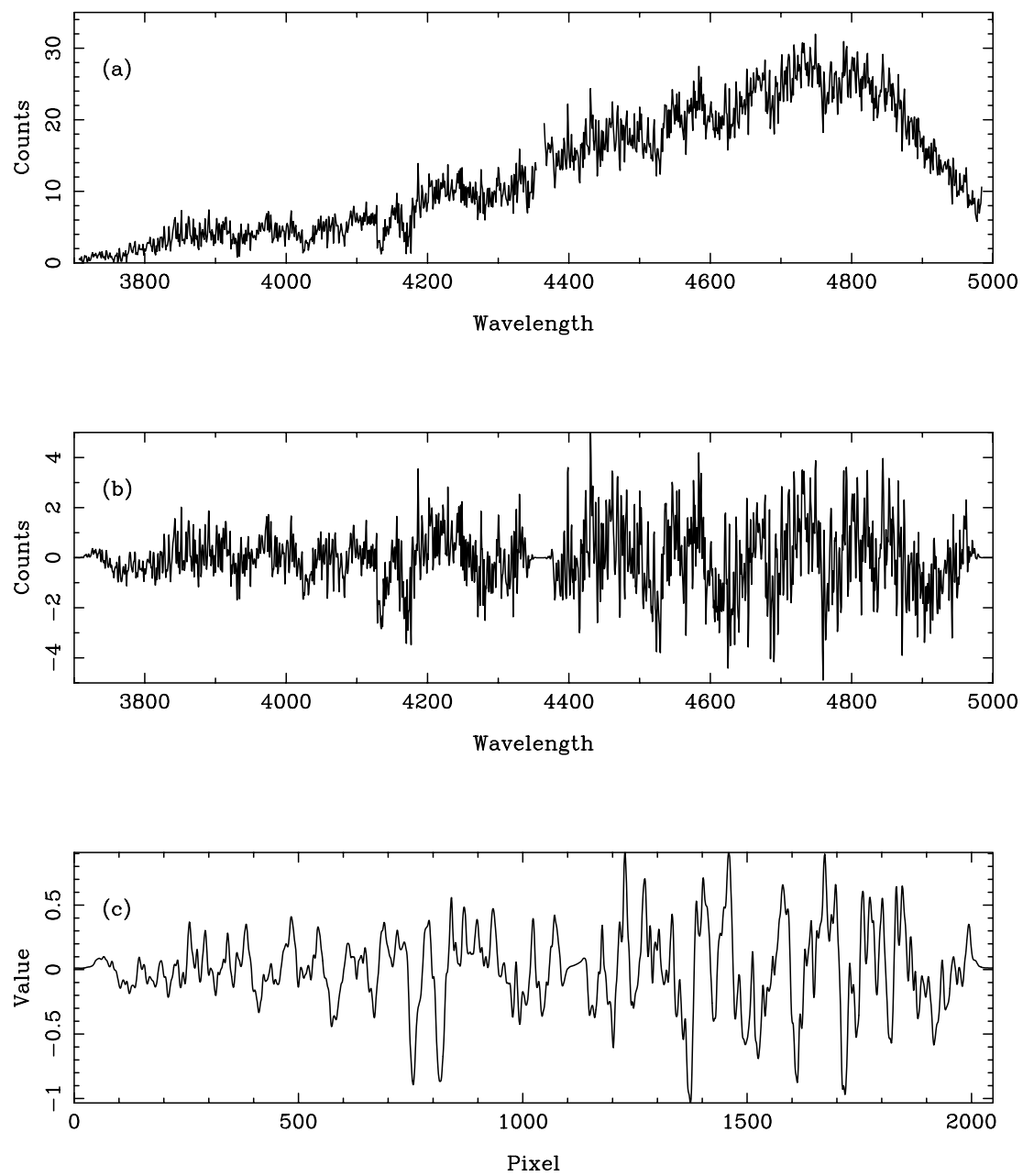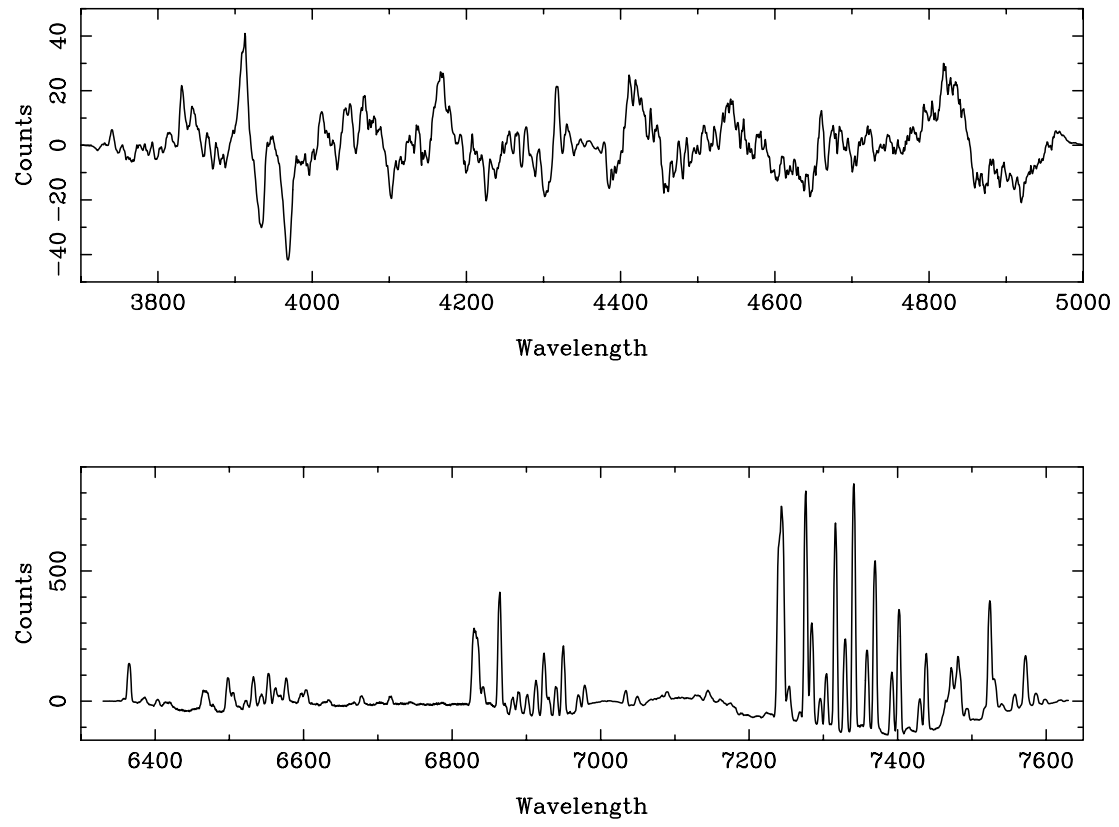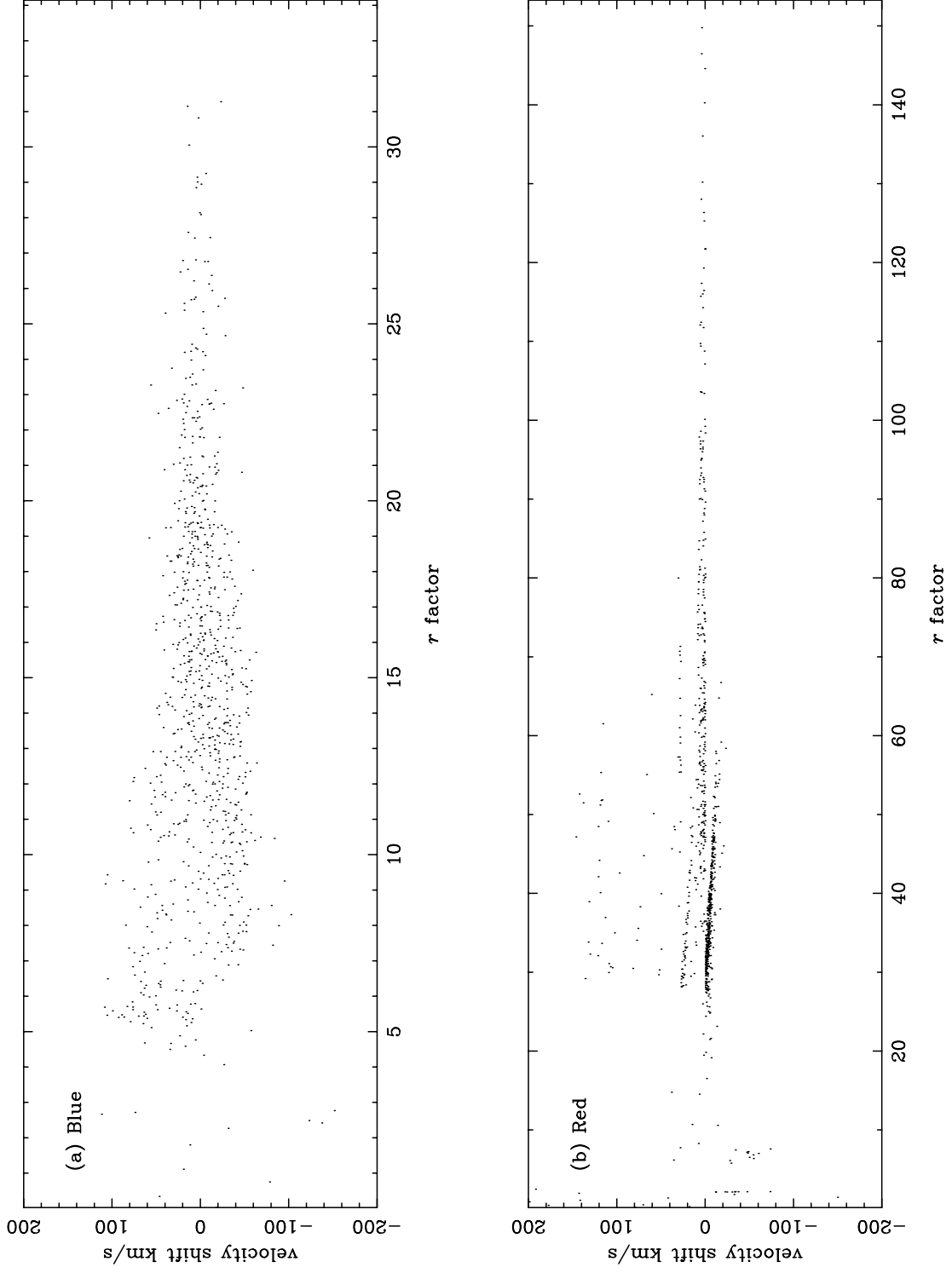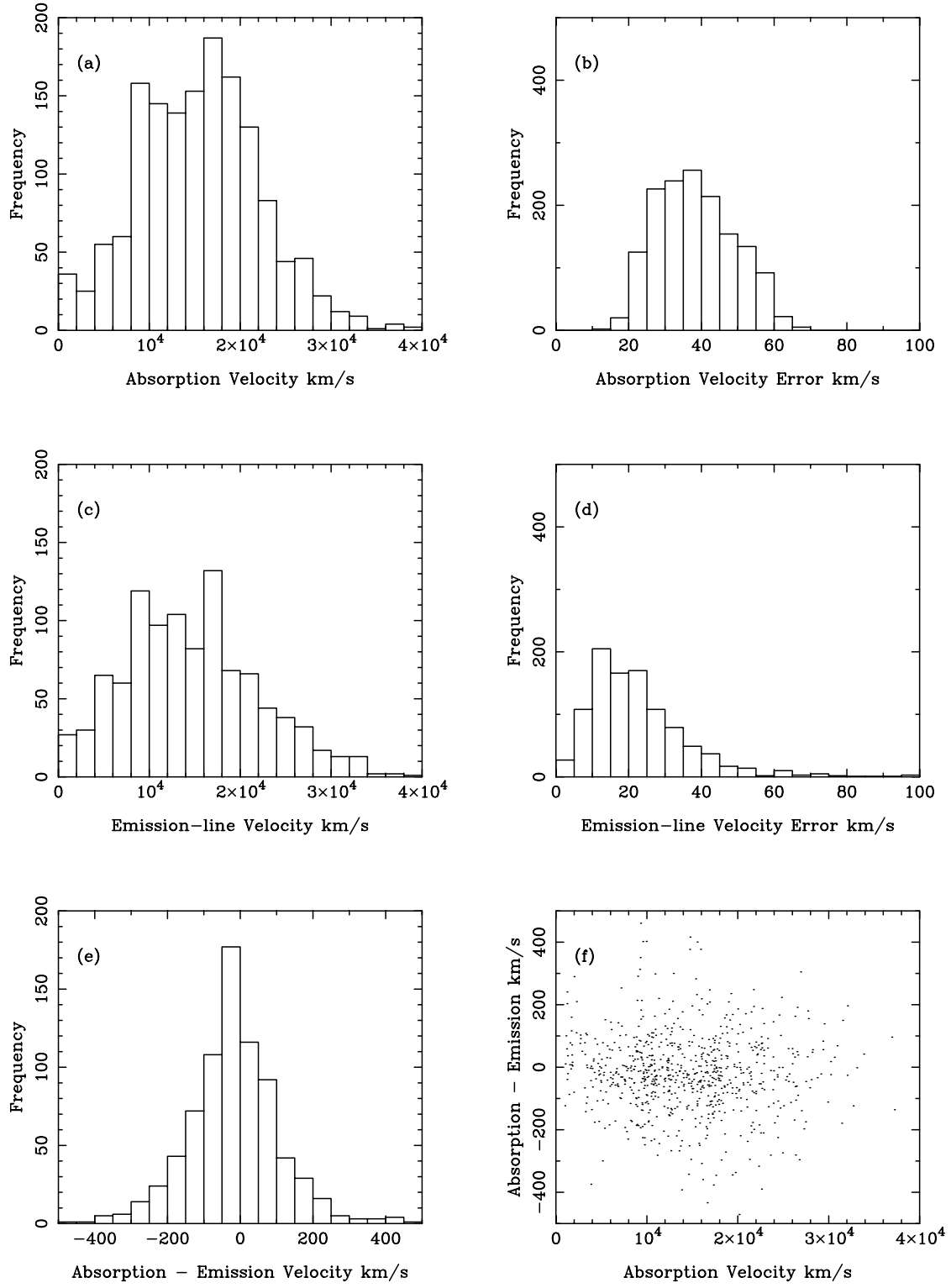
Fig. 2.—
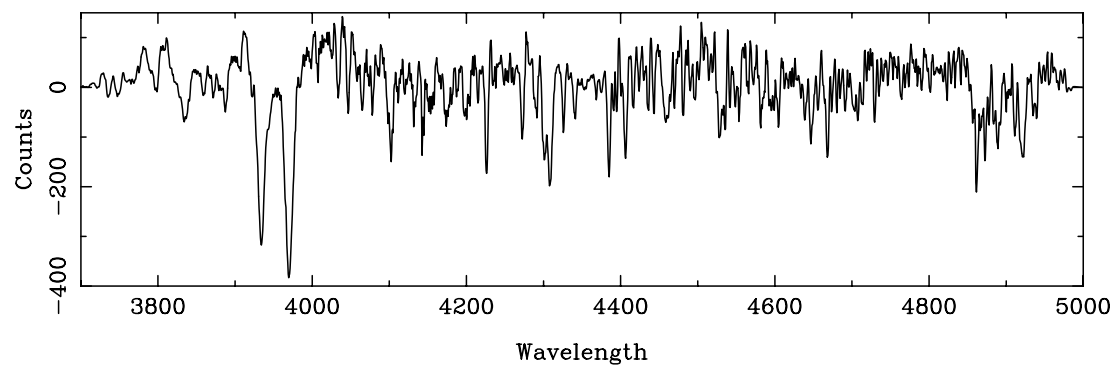
Fig. 3.—
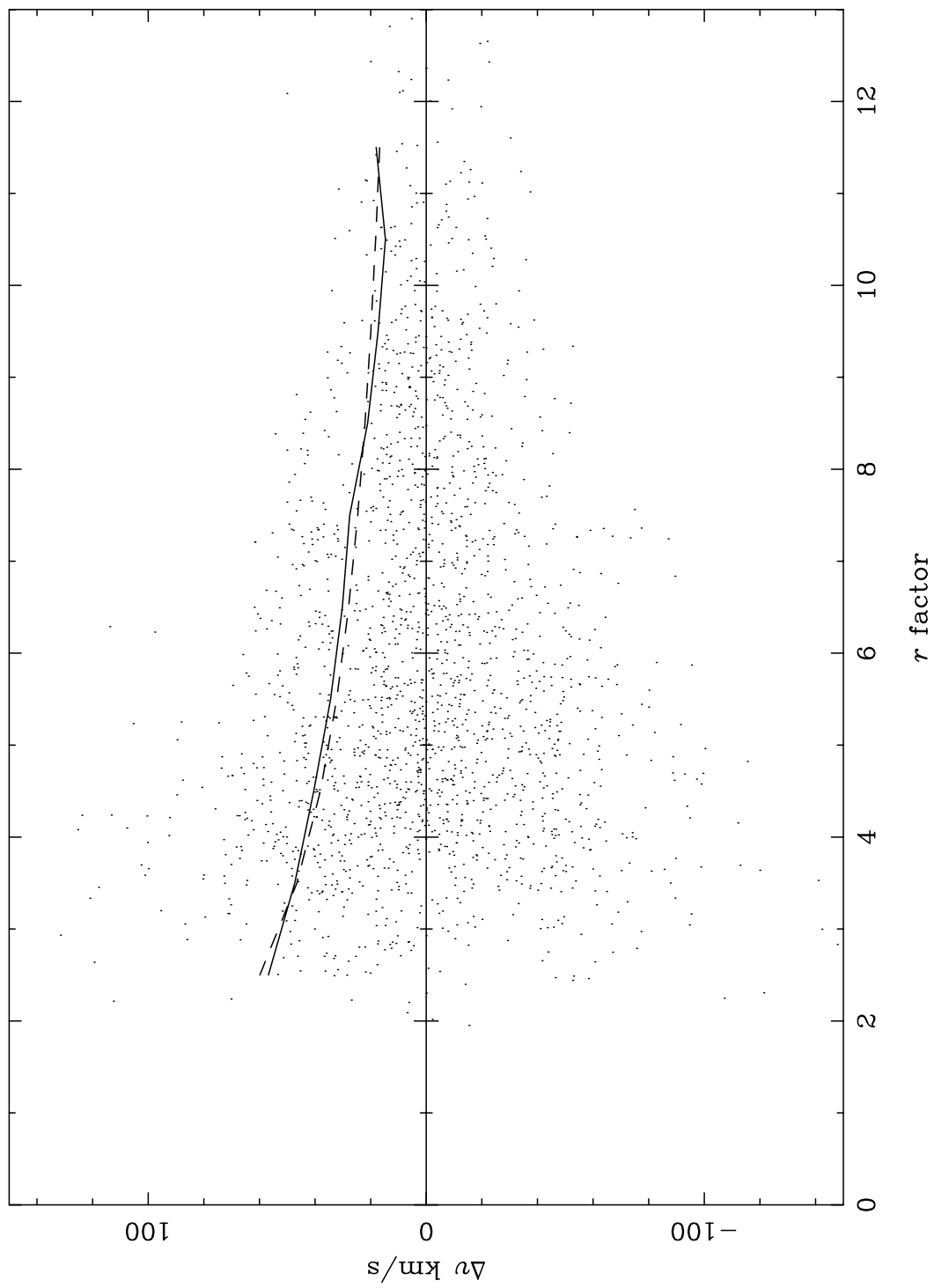
Fig. 4.—
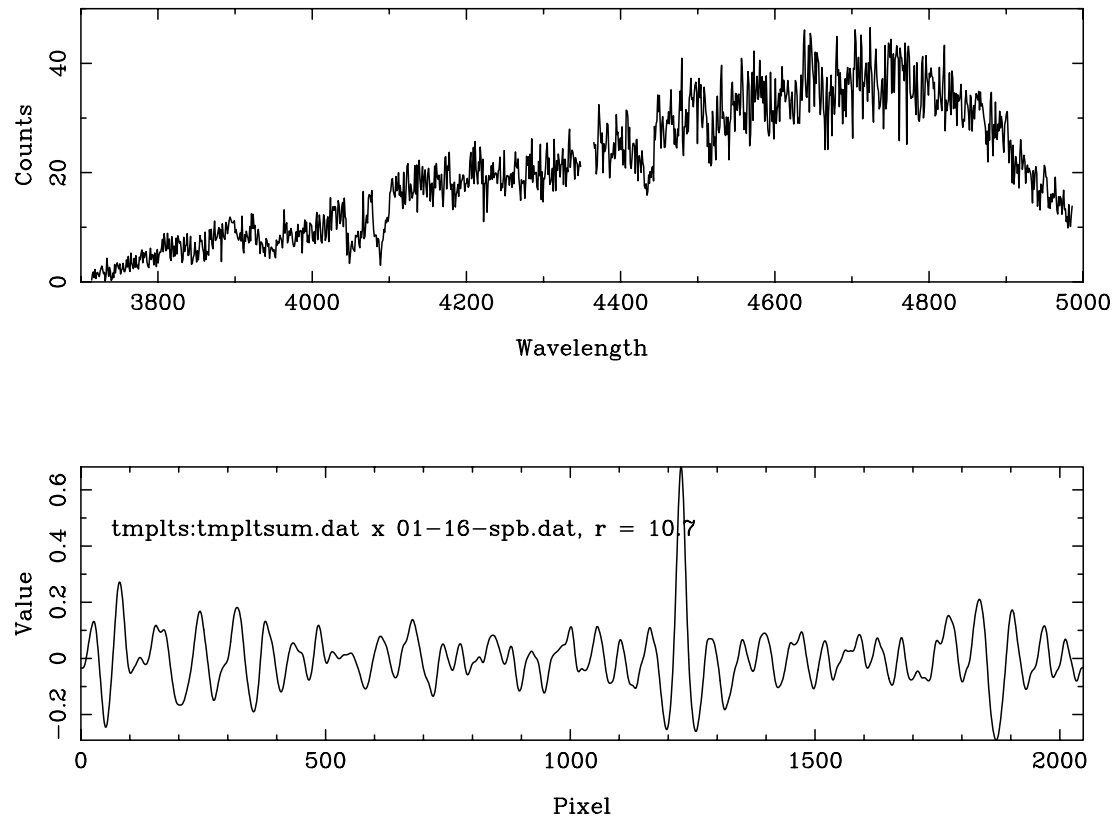
Fig. 5.—

Fig. 6.—

Fig. 7.—

Fig. 8.—

Fig. 9.—

Fig. 10.—

Fig. 11.—

Fig. 12.—

Fig. 13.—